Two-Dimensional Model of Hidden Markov Lattice

Bong-Kee Sin[†]

ABSTRACT

Although a number of variants of 2D HMM have been proposed in the literature, they are, in a word, too simple to model the variabilities of images for diverse classes of objects; they do not realize the modeling capability of the 1D HMM in 2D. Thus the author thinks they are poor substitutes for the HMM in 2D. The new model proposed in this paper is a hidden Markov lattice or, we can dare say, a 2D HMM with the causality of top-down and left-right direction. Then with the addition of a lattice constraint, the two algorithms for the evaluation of a model and the maximum likelihood estimation of model parameters are developed in the theoretical perspective. It is a more natural extension of the 1D HMM. The proposed method will provide a useful way of modeling highly variable patterns such as offline cursive characters.

이차원 은닉 마르코프 격자 모형

신 봉 기[†]

요 약

HMM이 시계열 모델로써 우수함이 널리 입증되면서 이차원 모델로 확장해 보려는 연구 결과도 늘어났지만 아직까지 임의의 객체 패턴의 다양한 변형을 모델링하기에는 너무 단순한 경우가 대부분이다. 따라서 HMM이 시계열 데이터에서 보여준 성과를 영상 데이터에서 기대하기는 어렵다. 즉, 아직 대부분의 모델이2D HMM으로 보기에는 부족하다고 판단된다. 본 논문에서 제안하는 모델은 이차원 공간에서 상하, 좌우 방향의 진행관계(causality)가 존재하는 은닉 마르코프 격자 또는 HML이다. 여기에 격자 구성 조건을 추가하여 모델 평가와 디코딩, 그리고 MLE 매개변수 추정법에 의한 훈련 알고리즘을 이론적으로 유도, 개발하였다. 본 모델은 기존의 필드형 모델과 달리 필기 문자 영상과 같이 다양한 국소적 형태 변형을 효과적으로 모델링하는 유용한 방법으로 사용될 수 있다.

1. Introduction

Hidden Markov model or HMM is a well-known statistical modeling tool for a variety of highly variable time-series or time-series-like signals. Its success, however, is limited to the analysis of 1D signals with an intrinsic order relation. Motivated by the success of the HMM, a number of researchers have tried to apply the model or its extensions to spatial 2D signals like digital images. The efforts thus far, however, have not been successful, although not a total failure either. This paper centers on the theoretical development of a 2D extension of the HMM.

The research on Markov models for 2D patterns is not new. It has been studied in several related areas such as image processing and character recognition. Although historically later to appear, the pseudo 2D HMM or P2DHMM is a simplified model of two-level hierarchy[1]; it is essentially a 1D HMM with vertical frame observations[2]. The

본 연구는 1999년도 부경대학교 기성회 연구비 지원으로 수행되었음.

[†] 부경대학교 전자컴퓨터정보통신공학부 전임강사

P2DHMM is cost-effective for modeling patterns free of global shape deformation. This is also generally true of the truly 2D model of Markov random field or MRF[3,4]. The MRF model has been studied and used by numerous researchers from diverse fields who are grappling with texture analysis, image restoration and segmentation [5]. It is, therefore, not strange that there are a large number of variants like Markov mesh and hidden Markov random field. For reasons of computational complexity most of the researchers adopted the causal types of MRFs for modeling images. Although there are studies on symmetric local dependence without directional causality, those methods suffer from cost-ineffectiveness problems and thus often resort to heuristic recipes. There is one study referring to 2D planar HMM[6]. The paper, however, focussed mainly on DP-based image match and made just a passing remark on the use of HMM without mathematical and/or practical development.

This paper describes a new development of 2D HMM. Distinct from the previous oversimplified Markov random field, the new model is called a hidden Markov lattice or HML that involves the causality of top-down and left-right direction. This causality, although not explicitly present in images, allows an efficient computation. In addition this paper introduces a lattice constraint under which HML can be locally scaled up or down just like the 1D time-warping of HMM. With the lattice constraint, the algorithms for evaluation of a model and maximum likelihood estimation of model parameters are developed.

The proposed model is different from the MRF in that homogeneity (and isotropism, of course) is not enforced. The effect is that many more interesting forms of image variations including local shape distortions can be modeled more systematically. In other words, with the HML, many more types of image distortions can now be explained with the most likely Markov lattice constructed by the decoding algorithm.

In the rest of the paper, we will first review the conventional HMM briefly in Section 2, and then address the definition of the HML and the DP-based evaluation algorithm in Section 3. The description of the algorithm is the 2D version of Viterbi algorithm. Therefore the model decoding in Section 4 will be made short with a few additional remarks. Section 5 presents an MLE-based re-estimation algorithm for training model parameters, which is the most difficult task of the three. The final section discusses implications of the proposed method, and then concludes the paper.

2. Hidden Markov Model

HMM is a statistical model for analyzing time sequential signal that can be viewed as one-dimensional. The time series data is characterized by a strict order that can be described by the time evolution of the model states. In the HMM theory, the evolution is modeled by an underlying Markov chain with probabilistic transitions between states. This is the first stochastic process of the HMM. The second process concerns the generation of symbols that can be observed outside, hence termed an observational process.

The conventional HMM is not of direct concern here. But a formal description may be helpful – or even required – for a clear and accelerated understanding of the theoretical aspect of HMM extensions. Let us first denote q_t, $t = 1, \ldots$ a stochastic process of Markov chain, and S the set of distinct states that the model takes. Formally the HMM is characterized by three sets of probabilistic parameters, namely :

- State transition probability, $A = \{a_{ij} : a_{ij} = p(q_{t+1} = j \mid q_t = i),\ i, j \in S\}$. This is the probability of changing states from i at time t into state j at time $t+1$. The parameters satisfy the stochastic constraint $a_{ij} \geq 0$ and $\sum_{j \in S} a_{ij} = 1$.
- Observation probability, $B = \{b_i(k) : b_i(k) =$

$p(x_t = k \mid q_t = i)\}$. This is the probability distribution function for symbol observation. These parameters are also subject to the constraint $\sum_{v \in V} b_i(v) = 1$ where V is the set of observable symbols in state i.

- Initial state transition probability, $\pi = \{\pi_i : \pi_i = p(q_1 = i)\}$. This is the probability of Markov chains starting at state i, and satisfies $\sum_{i \in S} \pi_i = 1$.

With these parameters HMMs can describe diverse patterns with a wide range of variability. The conventional HMM is often denoted as $\lambda = (A, B, \pi)$.

Since the first introduction of an efficient algorithm for estimating HMM parameters, several alternative estimation methods have been proposed for the purpose of improving the discrimination power or simplifying the calculation. But the Baum *et al.*'s method[7] has always been the basis of elaboration, and the same is true of the method to be described in this paper. The original proof of Baum-Welch algorithm dealt with specifically with a finite alphabet and general output distribution. A more generalized proof was based on constructing an information-theoretical Q-function, i.e., Kullback-Leibler number[8]. The same discussion will be given in Section 5, and it will not be duplicated here. For comparison to be made later in the section, the reestimation formula for transition parameters are given by

$$a'_{ij} = \sum_t P(\mathbf{X}, q_{t-1}=i, q_t=j \mid \lambda) / \sum_j \sum_t P(\mathbf{X}, q_{t-1}=i, q_t=j \mid \lambda).$$

It is almost of the same form as those found in Section 5, so the rest will not be given here. For more detailed description, refer to the paper by Rabiner[9].

3. Hidden Markov Lattice

3.1 Markov Lattice

A stochastic process $\mathbf{X} = \{X_n, n = 1, 2, \ldots\}$ where each variable taking on a finite number of possible values is a Markov chain if there is a fixed probability

$$P\{X_{n+1}=j \mid X_n=i, X_{n-1}=i_{n-1}, \ldots, X_1=i_1, X_0=i_0\} = P_{ij}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and $n \geq 0$. This type of Markov chain is sufficient for modeling one-dimensional time-series signals where a variable is related to only one or less variable that is preceding it when viewed in time dimension. In higher dimensions, however, this type of simple chain structure is not adequate, and one or more additional variables are required. Let us, from now on, limit our discussion to two-dimensional signals such as a rectangular image consisting of a lattice of pixels. Needless to say, the model will be easily extended to third or higher dimensions.

There are two equivalent ways of defining random configurations of points on a lattice. One is based on the formulation of statistical mechanics according to J. Gibbs. Called as Gibbs ensemble or Gibbs random field, it is generally accepted as the simplest useful mathematical model of discrete or lattice gas. The second class of random fields is Markov random field, whose foundation dates back to the physics literature on ferromagnetism originating in the work of E. Ising in 1925[10]. This extends in a simple way the notion of Markov process with one dimensional, integer valued, time to the case of higher dimensional, lattice valued, space parameter.

Let $\mathbf{L} = \{(i, j) : 1 \leq i \leq M, 1 \leq j \leq M\}$ be a two-dimensional rectangular lattice with $L = MN$ sites arranged as a planar mesh. M and N denote the vertical and horizontal dimension of the lattice respectively. For convenience let us denote the state or site identifiers as $i = 1, \ldots, L$ in row-major order. For each site i in the lattice we find a set of sites which are adjacent to and thus condition the state of the current site. It is called a neighborhood. The neighborhood of a site i in the lattice $\mathbf{L}$ is a set of sites that can influence behavior of the site i. In general the neighborhood system is defined as follows : $\eta = \{\eta_i \subseteq \mathbf{L} : i \in \mathbf{L}\}$. Here η_i

is the neighborhood of a site i, and satisfies that $i \notin \eta_i$ and $j \in \eta_i$ if and only if $i \in \eta_j$. Then the definition of the MRF follows: given a lattice $\mathbf{L}$ and a neighborhood η, a random field $\mathbf{X} = \{X_j, j \in \mathbf{L}\}$ is an MRF if and only if

$$P(X_j = x_j | X_i = x_i, i \in \mathbf{L} - \{j\}) = P(X_j = x_j | X_i = x_i, i \in \eta_j), \forall j \in \mathbf{L}.$$

By definition, the MRF is homogeneous and isotropic. This property is highly appropriate for modeling systems of homogeneous gas particles or fluids, and restoring images corrupted by random noise. But the problem of such a noncausal random field model is that there is no known efficient and effective algorithm other than the formulation based on the Gibbs distribution. Several researchers have tried to solve the problem by introducing causality in the lattice. Two recent studies were reported by Park *et al.* [11,12].

However, the MRF is still insufficient for modeling general image distortions other than random corruption of images. There are many more types of characteristic variations of images arising not from purely random sources but from sources explainable in statistical terms. This is particularly true of hand-written script. Such images involve local distortions characteristic of the target patterns in the images. We believe that they should be modeled with a new type of modeling framework that can represent various local variations. One proposed in this paper is based on the model of Markov lattice.

Just like an MRF, a general Markov lattice has it that a site is determined by a set of neighbor sites. The difference is lies in the definition of anisotropic inhomogeneous click potential which is defined as a probabilistic transition parameter

$$P(X_j = x_j | X_i = x_i, i \in \eta_j) = \Pi_{i \in \eta_j} P_{ij} \quad \forall j \in \mathbf{L}$$

satisfying the stochastic constraint

$$P_{ij} \geq 0,$$
$$\Sigma_{j \in \eta_i} P_{ij} = 1.$$

It is not necessarily that $P_{ij} = P_{ji}$ and $P_{ij} = P_{i+k,j+k}$, and this property allows the modeling of local spatial distortion.

3.2 2D Hidden Markov Lattice

Based on the concept of the Markov lattice of the preceding section and the traditional HMM theory, we can define a two-dimensional hidden Markov lattice (HML). The model to be described henceforth is causal and allows an efficient computation. Formally the HML is defined as follows :

- Site transition parameters

In a multi-dimensional space free of temporal arrow it is difficult to justify the introduction of any order, or causality. However, we have assumed an intuitive causality to reduce the computational requirement in the following way: first, there are two types of transitions: the downward transition from the upper neighborhood and the rightward from the left neighborhood. Second, we restrict the site transitions to those to and from the set of 8-neighbors.

The resulting mesh topology of the model is shown in Figure 1. For a given node, say j, the set of upper neighbors that can act as an upper source node is denoted by η_j^U. Similarly the left neighborhood is denoted by η_j^L. The right and the lower neighborhoods η_j^R and η_j^D respectively denote the sets of right and the lower destination nodes of rightward and downward transitions respectively from the site j. Using the two types of transitions between neighbor sites, we can construct a complete lattice $\mathbf{Y}$ of 2D HML states corresponding to an image, a lattice of pixels.

In 2D space there are two types of transitions: vertical downward transitions and horizontal rightward transitions, each parameterized as follows :

$$a_{hj}^{\downarrow} = p(q_u = j | q_{u-V} = h),\ h \in \eta_j^U = \{j-N-1, j-N, j-N+1, j\},$$
$$a_{ij}^{\rightarrow} = p(q_u = j | q_{u-1} = i),\ i \in \eta_j^L = \{j-N-1, j-1, j+N-1, j\}$$
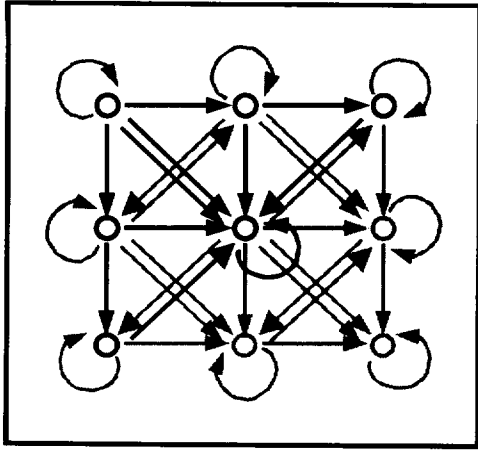
Figure 1. HML mesh topology.

These parameters define the causality, vertical and downward, as assumed before. Also it is noted that the following stochastic constraints are satisfied

$$\Sigma_{k\in\eta_j^R}\, a_{jk}^{\rightarrow} = 1, \qquad a_{jk}^{\rightarrow} \geq 0,$$
$$\Sigma_{l\in\eta_j^D}\, a_{jl}^{\downarrow} = 1, \qquad a_{jl}^{\downarrow} \geq 0 \qquad (1)$$

Here η_j^R and η_j^D denote the right and the lower neighborhood respectively.

- Observation parameters

In the conventional HMM, the observation is another stochastic process which is a probabilistic function of an underlying Markov chain. The observation of an HML is an image $\mathbf{X} = \{x_{uv} \in \Omega: 1 \leq u \leq U, 1 \leq v \leq V\}$ in the rectangular arrangement of $W = UV$ pixels. $\Omega = \{1, 2, ..., K\}$ is a set of K color or gray scale values. Here again let us identify the pixels in the row-major order as $u = 1, ..., W$.

The observation of of $\mathbf{X}$ is a function of the above lattice process parameters. The observation symbols x_{uv} are independent of all the others. This type of conditional independence assumption is grossly inaccurate, but allows an efficient computation and usually works well enough. The parameters are :

$$b_j(v) = P(x_u = v | q_u = j), \qquad j \in \mathbf{L},\ v \in \Omega,$$

where

$$\Sigma_v\, b_j(v) = 1, \qquad j \in \mathrm{L}.$$

Every site in an HML is conditioned by its neighboring sites, and the collection of the sites organizes a lattice through an artificial causal chain. Each pixel in an image $\mathbf{X}$ is observed from a HML site as a result of a conditionally independent process of the corresponding site. The capability of modeling spatial distortions or spectral variations depends on the organization of neighborhood system (as in the case of MRF) or the transition probability parameters of the 2D HML.

3.3 Lattice Process

Now, given the causality for a 2D lattice, we can proceed to define the following two recurrence relations based on the Markovian property and the Bellmans optimality principle of dynamic programming. They are the forward probability and the backward probability:

$$\alpha_u(j) = \max_{h\in\eta_j^U,\ i\in\eta_j^L} a_{ij}^{\rightarrow} a_{hj}^{\downarrow}\ b_j(x_u)\ \alpha_{u-1}(i),$$
$$j = 1, ..., L, \quad u = 1, ..., W \qquad (2)$$

$$\beta_u(j) = \max_{k\in\eta_j^R,\ l\in\eta_j^D} a_{jk}^{\rightarrow} a_{jl}^{\downarrow}\ b_k(x_{u+1})\ \beta_{u+1}(k),$$
$$j = L, ..., 1, \quad u = \mathrm{W}, ..., 1. \qquad (3)$$

The forward probability $\alpha_u(j)$ is the maximum probability of observing the partial region of the image $\mathbf{X}_{1,u} = x_1\, x_2 \ldots x_u$ from the partial mesh of states $\mathbf{Y}_{1,j} = y_1\, y_2 \ldots y_j$. The backward probability $\beta_u(j)$ denotes the probability of observing the remaining image region $\mathbf{X}_{u+1,W} = x_{u+1}\, x_{u+2} \ldots x_W$ after x_u from the remaining partial mesh of states $\mathbf{Y}_{j+1,L} = y_{j+1}\, y_{j+2} \ldots y_{Ls}$. The boundary conditions are :

$$\alpha_1(1) = b_1(x_1)$$
$$\alpha_u(j) = \max_{i\in\eta_j^L} a_{ij}^{\rightarrow} b_j(x_u)\ \alpha_{u-1}(i), \quad j = 1, ..., N,$$
$$u = 2, ..., V$$

$$\beta_W(L) = 1$$
$$\beta_u(j) = \max_{k\in\eta_j^R} a_{jk}^{\rightarrow} b_k(x_{u+1})\ \beta_{u+1}(k),$$
$$j = L, ..., L-N+1,\ u = W, ..., W-V+1.$$

The forward DP continues while keeping the mesh lattice-related information as

$$(h^*, i^*)_u(j) = \text{argmax}_{h \in \eta_j^U, i \in \eta_j^L} a_{ij}^{\rightarrow} a_{hj}^{\downarrow} b_j(x_u) \times \alpha_{u-1}(i), \quad j = i, \ldots, L, \; u = 1, \ldots, W \quad (4a)$$

$$(k^*, l^*)_u(j) = \text{argmax}_{k \in \eta_j^R, l \in \eta_j^D} a_{jk}^{\rightarrow} a_{jl}^{\downarrow} b_k(x_{u+1}) \times \beta_{u+1}(k), \quad j = L, \ldots, 1, \; u = W, \ldots, 1 \quad (4b)$$

Here, let us write

$i^* = Left(j)$
$h^* = Up(j)$
$k^* = Right(j)$
$l^* = Down(j)$

Then we have another requirements for building a complete mesh lattice of states as the result of computation in (2) and (3).

$$Left(h^*) = Up(i^*), \quad Down(k^*) = Right(l^*) \quad (5)$$

and

$$\alpha_u(j) = \max_{h \in \eta_j^U} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(R_h), \quad j = 1, N+1, \ldots, (M-1)N+1, \; u = V+1, 2V+1, \ldots, (U-1)V+1, \quad (6a)$$

$$\beta_u(j) = \max_{l \in \eta_j^D} a_{jl}^{\downarrow} b_k(x_{u+1}) \beta_{u+1}(L_l), \quad j = L, L-N, \ldots, N, \; u = W-V, W-2V, \ldots, V. \quad (6b)$$

Here $R_h = Right^{N-1}(h)$, $L_l = Left^{N-1}(l)$, and each indicates the rightmost boundary node of h and the leftmost boundary node of l. The power notation is defined by the recursion for all n as a composite function:

$$Right^{n}(x) = Right(Right^{n-1}(x))$$
$$Left^{n}(x) = Left(Left^{n-1}(x)).$$

The above equations (5) and (6) constitute the lattice constraints for a complete mesh lattice.

Using the forward and backward probabilities of (2) and (3), we can complete the calculation as

$$P(\mathbf{X}|\Lambda) = \max_j \alpha_u(j) \left[\prod_{k=1}^{V-1} a_{h_k i_k}^{\downarrow}\right] \beta_u(j) \quad (7)$$

Here again any complete mesh lattice requires that

$h_k = y_{u-V+k} = Right(h_{k-1})$ and $h_0 = h$,
$j_k = y_{u+k} = Right(j_{k-1})$ and $j_0 = j$,

the condition for sewing together the forward and the backward lattice patches to obtain a complete rectangular lattice. The resulting mesh of states will be a planar lattice locally warped to model the locally deformed 2D patterns. And it is noted that this type of two-dimensional lattice model is different from the second order Markov chain[13] in that this does not impose the lattice constraint which leads to the construction of a regular mesh of states.

4. Lattice Decoding

Apart from patterned texture or random noise modeling, an observed image **X**, in general, does not possess the Markovian property, and thus it can not be modeled properly with an MRF. In this paper an image is defined as a realization of a stochastic process that, in turn, is defined over an underlying Markov Lattice. Each observation is a function of the corresponding site of the Markov lattice but is assumed to be independent of other observations. The individual site states of a Markov lattice are determined based on their neighborhood. But, since the states in the lattice depend on their location, the model is also distinguished from the homogeneous MRF that is not aware of model topology.

The decoding of 2D HML Λ is the problem of finding the optimal Markov lattice $\mathbf{Y}^*$ of maximum likelihood given an observation image **X**. Mathematically it is defined as the task of maximizing $P(\mathbf{X},\mathbf{Y}|\Lambda)$ over all possible chains **Y**. **Y** is a complete mesh lattice of sites. In the preceding section, we have already defined the forward probability in (2) in terms of the best realization of initial partial lattices. The final probability is the very result of

decoding. Finally the optimal Markov mesh lattice can be obtained by backtracking the result of forward pass using the information of (4) after the forward pass is over.

In standard theory of hidden Markov modeling, the model evaluation is based on the concept of total probability of observing an input signal given a model, which is given by

$$P(\mathbf{X}|\Lambda) = \Sigma_{\mathrm{L}}\ P(\mathbf{X},\mathbf{L}|\Lambda)\ P(\mathbf{L}|\Lambda)$$

Although correct in statistical context, there is a difficulty in interpreting the result of computation. Namely, given a model of planar topology, one is asked whether it is possible to generate the rectangular image without regard to the topology. This problem leads us to define the optimization criterion as the joint probability of the lattice of states as well as the input image. Formally it is given by the following formula

$$P(\mathbf{X}|\Lambda) = \max_j\ \alpha_u(j)\ [\Pi_{k=1}^{V-1}\ a_{hk\ ik}^{\downarrow}]\ \beta_u(j)$$

as is given in the preceding section. In effect, this is the equation for decoding the model Λ given an image $\mathbf{X}$.

5. Parameter Estimation

The parameter estimation problem is concerned with finding the optimal set of model parameters given a set of typical samples. Let us write $\mathbf{Y}$ be a Markov mesh chain given a sample image $\mathbf{X}$. The likelihood of observing $\mathbf{X}$ from the Λ of the model is

$$P(\boldsymbol{X}|\Lambda) = \Sigma_{\mathbf{Y}}\ P(\mathbf{X},\mathbf{Y}|\Lambda).$$

Each term in the right hand side is the joint probability written as

$$P(\mathbf{X},\mathbf{Y}|\Lambda) = \Pi_{u=1}^{W}\ [a_{yu-1,yu}^{\rightarrow}\ a_{yu-V,yu}^{\downarrow}\ b_{yu}(x_u)]. \quad (8)$$

By taking the logarithm of it, we have

$$\log P(\mathbf{X},\mathbf{Y}|\Lambda) = \Sigma_{u=1}^{W}\ (\log a_{yu-1,yu}^{\rightarrow} + \log a_{yu-V,yu}^{\downarrow} + \log b_{yu}(x_u)) \quad (9)$$

Following the line of Baum's reasoning with the Q-function[7], we can now define a similar auxiliary for the 2D HMM as follows:

$$\begin{aligned} Q(\Lambda,\ \Lambda') &= 1/P(\mathbf{X}|\Lambda)\ \Sigma_{\mathbf{Y}}\ P(\mathbf{X},\ \mathbf{Y}|\Lambda)\ \log P(\mathbf{X},\ \mathbf{Y}|\Lambda') \\ &= 1/P\ \Sigma_{\mathbf{Y}=y1\ y2\ \ldots\ yW}\ P(\mathbf{X},\ \mathbf{Y}|\Lambda) \times \Sigma_{u=1}^{W}\ (\log a'^{\rightarrow} + \log a'^{\downarrow} + \log b') \\ &= 1/P\ \Sigma_i\ \Sigma_j\ \Sigma_u\ P(\mathbf{X},\ y_{u-1}=i,\ y_u=j|\Lambda)\ \log a'^{\rightarrow} \\ &\quad + 1/P\ \Sigma_h\ \Sigma_j\ \Sigma_u\ P(\mathbf{X},\ y_{u-V}=h,\ y_u=j|\Lambda)\ \log a'^{\downarrow} \\ &\quad + 1/P\ \Sigma_j\ \Sigma_k\ \Sigma_u\ P(\mathbf{X},\ y_u=j|\Lambda)\ \log b' \end{aligned} \quad (10)$$

where $P = P(\mathbf{X}|\Lambda)$. The last expression can be reorganized into

$$\mathrm{Q}(\Lambda,\ \Lambda') = \Sigma_h\ \Sigma_j\ c_{hj}\ \log a'^{\downarrow}_{hj} + \Sigma_i\ \Sigma_j\ d_{ij}\ \log a'^{\rightarrow}_{iji} + \Sigma_j\ \Sigma_k\ e_{jk}\ \log b_j'(x_k) \quad (11)$$

where

$$\begin{aligned} c_{hj} &= \Sigma_u\ P(\mathbf{X},\ y_{u-1}=i,\ y_u=j|\Lambda)\ /\ P(\mathbf{X}|\Lambda) \\ d_{ij} &= \Sigma_u\ P(\mathbf{X},\ y_{u-V}=h,\ y_u=j|\Lambda)\ /\ P(\mathbf{X}|\Lambda) \\ e_{ik} &= \Sigma_{u:xu=k}\ P(\mathbf{X},\ y_u=j|\Lambda)\ /\ \Sigma_u\ P(\mathbf{X},\ y_u=j|\Lambda). \end{aligned}$$

Then the resulting formulae for re-estimating the parameters are as follows:

$$a'^{\rightarrow}_{ij} = \Sigma_u\ P(\mathbf{X},y_{u-1}=i,y_u=j|\Lambda)\ /\ \Sigma_j\ \Sigma_u\ P(\mathbf{X},y_{u-1}=i,\ y_u=j|\Lambda) \quad (12)$$

$$a'^{\downarrow}_{hj} = \Sigma_u\ P(\mathbf{X},y_{u-V}=h,y_u=j|\Lambda)\ /\ \Sigma_j\ \Sigma_u\ P(\mathbf{X},y_{u-V}=h,\ y_u=j|\Lambda) \quad (13)$$

$$\begin{aligned} b'_{jk} &= \Sigma_{u:xu=k}\ P(\mathbf{X},y_u=j|\Lambda)\ /\ \Sigma_k\ \Sigma_{u:xu=k}\ P(\mathbf{X},y_u=j|\Lambda) \\ &= \Sigma_{u:xu=k}\ P(\mathbf{X},y_u=j|\Lambda)\ /\ \Sigma_u\ P(\mathbf{X},y_u=j|\Lambda) \end{aligned} \quad (14)$$

Let us consider the Q-function of EM algorithm as a function of Λ'. Although the above function has more parameters than the corresponding function of 1D HMM, they are essentially of the same form. Therefore we can say the above re-estimation algorithm converges. It is stated in the following theorem.

[Theorem 1] *If* $Q(\Lambda,\ \Lambda') \geq Q(\Lambda,\ \Lambda')$, *then* $P(\mathbf{X}|\Lambda') \geq P(\mathbf{X}|\Lambda)$. *The equality holds when* $P(\mathbf{X}|\Lambda') = P(\mathbf{X}|\Lambda)$.

Proof: From the concavity of the log function

it follows that

$$\begin{aligned}\log P(\mathbf{X}|\Lambda')/P(\mathbf{X}|\Lambda) &= \log [\Sigma_{\mathbf{Y}} P(\mathbf{X},\mathbf{Y}|\Lambda')/P(\mathbf{X}|\Lambda)] \\ &= \log [\Sigma_{\mathbf{Y}} P(\mathbf{X},\mathbf{Y}|\Lambda)/P(\mathbf{X}|\Lambda) \times P(\mathbf{X},\mathbf{Y}|\Lambda')/P(\mathbf{X},\mathbf{Y}|\Lambda)] \\ &\geq \Sigma_{\mathbf{Y}} P(\mathbf{X},\mathbf{Y}|\Lambda)/P(\mathbf{X}|\Lambda) \times \log [P(\mathbf{X},\mathbf{Y}|\Lambda')/P(\mathbf{X},\mathbf{Y}|\Lambda)] \\ &= Q(\Lambda, \Lambda') - Q(\Lambda, \Lambda)\end{aligned}$$

where the inequality is due to the well-known Jensens inequality. The above inequality says that Λ is a critical point of $P(\mathbf{X}|\Lambda)$ if and only if it is a critical point of Q as a function of Λ'. QED.

According to the above result, if the newly estimated model Λ' makes the right-hand side positive, the algorithm is guaranteed to improve the model likelihood $P(\mathbf{X}|\Lambda')$. The improvement then results in Λ' that maximizes the Q-function unless a critical point is reached[7].

6. Discussion and Conclusion

The most essential difference between 1D time series data and 2D image is the existence or absence of intrinsic order between components in the signals. Thus we have raised the issue of introducing a putative order in order to utilize the sequential processing of modern computers.

The MRF as a model for 2D image is described by a small number of parameters defining clique potential (instead of transition probability) subject to the global homogeneity condition. Unlike the MRF or Markov mesh field models, the 2D HML of the paper was defined based on the concept of the probabilistic function of Markov mesh lattice. The lattice is a natural extension of the sequential chain. Therefore the HML is a natural extension 1D HMM, and a truer 2D HMM.

The causality referred to in the paper is not new; it has already been used in the previous studies on mesh models such as mesh random field. One distinguishing feature of the current method is the lattice constraint that constrains the search for only complete lattices. Naturally this has led to the use of Viterbi-type of decoding algorithm. Another noteworthy feature is that, with the introduction of site-to-site transition parameters, local spatial distortions can be modeled. We believe that this type of capability should be considered in modeling patterns of high variability, which is highly unpredictable globally, and thus less likely to be parameterized, but can be anticipated and modeled locally, remotely based on the study the psychomotor of handwriting. The author believes that the Markovian assumption fits appropriately with the latter points.

In Section 3 we have assumed a model topology with 2^{nd} order neighborhood in addition to directional causality. This will enable us to reduce the computational load drastically from $O(L^2W) = O(M^2N^2UV)$ of general ergodic models down to $O(LW) = O(MNUV)$ without decreasing the modeling power in general.

The 2D HML proposed in this paper is different from HMMRFs or mesh MRFs, and it is true even in the basic assumptions. The previous field models are too simple to model a wide range of shape variations occurring in images. But the HML is capable of decoding strictly local shape deformation, a task that may be called as dynamic space warping, in contrast to the dynamic time warping. The structure of the model is better suited for local nonlinear variation of a reference image, be it scaling or distortion either globally or locally.

The final remark is that the HML is not limited to two-dimensional space modeling, therefore the model can be called simply as HML instead of 2D HML. For 2D case, however, it is certain that the model will make a useful tool in such tasks as off-line handwritten character recognition, nonlinear motion field analysis.

References

[1] O. Agazzi and S. Kuo, "Pseudo two-dimen-

sional hidden Markov models for document recognition," *AT&T Technical Journal*, vol. 72, no.5, pp. 60-72, Sep.-Oct. 1993.
[2] W. Cho, S.-W. Lee and J. H. Kim, "Modeling and recognition of cursive words with hidden Markov models," *Pattern Recognition*, vol. 28, no. 12, pp. 1941-1953, 1995.
[3] J. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Info. Theory*, vol. IT-18, no. 2, pp. 232-240, March 1972.
[4] A. Jain, "Advances in mathematical models for image processing," *Proc. IEEE*, vol. 69, pp. 502-528, 1981.
[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.
[6] E. Levin and R. Pieraccini, "Dynamic planar warping for optical character recognition," in *Proc. Int'l Conf. On Accoustics, Speech and Signal Processing*, San Francisco, California, pp. 23-26, March 1992.
[7] L. Baum and J. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a models for ecology," *Bull. Ann. Math. Stat.*, vol. 73, pp. 360-363, 1967.
[8] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, pp. 79-86, 1951.
[9] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-285, 1989.
[10] E. Ising, *Zeitschrift Physik*, vol. 31, p. 253, 1925.
[11] H.-S. Park and S.-W. Lee, "A truly 2-D hidden Markov model for off-line handwritten character recognition," *Pattern Recognition*, vol. 31, no. 12, pp. 1849-1864, 1998.
[12] G. Saon, "Cursive word recognition using a random field based hidden Markov model," *Intl J. Doc. Aanal. and Recog.*, vol. 1, no. 4, pp. 199-208, 1999.
[13] A. Kundu, Y. He, and P. Bahl, "Recognition of handwritten word: first and second hidden Markov model based approach," *Pattern Recognition*, vol. 22, no. 3, pp. 283-297, 1989.

신 봉 기

1985년 서울대학교 자원공학 학사
1987년 한국과학기술원 전산학 석사
1995년 한국과학기술원 전산학 박사
1987년~1999년 한국통신 멀티미디어연구소
1999년~현재 부경대학교 전자컴퓨터정보통신공학부 전임강사
관심분야 : 인공지능, 패턴인식, 지능형 에이전트, 정보검색